

Optimal data collection in wireless sensor networks with correlated energy harvesting

Kishor Patil · Koen De Turck · Dieter Fiems

Received: date / Accepted: date

Abstract We study the optimal data collection rate in a hybrid wireless sensor network where sensor data is collected by mobile sinks. In such networks, there is a trade-off between the cost of data collection and the timeliness of the data. We further assume that the sensor node under study harvests its energy from its environment. Such energy harvesting sensors ideally operate energy neutral, meaning that they can harvest the necessary energy to sense and transmit data, and have on-board rechargeable batteries to level out energy harvesting fluctuations. Even with batteries, fluctuations in energy harvesting can considerably affect performance, as it is not at all unlikely that a sensor node runs out of energy, and is neither able to sense nor to transmit data. The energy harvesting process also influences the cost vs. timeliness trade-off as additional data collection requires additional energy as well. To study this trade-off, we propose an analytic model for the value of the information that a sensor node brings to decision making. We account for the timeliness of the data by discounting the value of the information at the sensor over time, we adopt the energy-chunk approach (i.e. discretise the energy level) to track energy harvesting and expenditure over time, and introduce correlation in the energy harvesting process to study its influence on the optimal collection rate.

Keywords Value of Information · Correlated energy harvesting · Markov Process · Sensor networks

Kishor Patil
Ghent University, Department of Telecommunications and Information Processing.
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium.
E-mail: patil.kishor@ugent.be

Koen De Turck
Central Supélec, Laboratoire des Signaux et Systèmes.
3, Rue Joliot-Curie, 91192 Gif-sur-Yvette, France

Dieter Fiems
Ghent University, Department of Telecommunications and Information Processing.
Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium

1 Introduction

Wireless sensor networks (WSNs) are one of the key enabling technologies of the Internet of Things (IoT) [1, 2], and have attracted considerable research interest over the past couple of years. WSNs are formed by a few to several thousands of interconnected devices called sensor nodes (SN) that monitor spatially distributed data like temperature, humidity, movement and noise [3, 4], extract information from the collected data and deliver relevant information to the user. WSNs have a variety of applications including military, environmental, home, industrial monitoring and healthcare. Applications of WSNs are surveyed in [5], whereas [6] focuses on various applications of WSNs in the context of the IoT.

The sensor node is the main component of a WSN. It's a small electronic device, powered by an on-board battery which can sense and transmit the data. For applications like remote sensing, it is often unfeasible or at least very expensive to replace the battery once a sensor node is deployed. Therefore a sensor's lifetime is largely determined by its energy budget on deployment. To increase the lifetime, one can opt to increase the battery size, but this incurs an additional cost and increases the weight and size of the SN which makes this solution less attractive [7]. An alternative approach focuses on optimising the energy expenditure of the SN. If less energy is used over time, the SN remains operative for a longer time with the same initial energy budget, see [8, 9]. Controlling the communication subsystem of the SN is a key component in the reduction of energy consumption [10]. The control can include duty cycling [11] or be based on computing the redundancy of nodes in the WSN [12].

Even if energy consumption is optimised, the lifetime of a SN is still determined by the limited energy budget on deployment. To overcome this dependence, one promising solution proposes networks of sensor nodes that can scavenge the needed energy from their environment, see [13]. Such sensor nodes are referred to as energy harvesting sensor nodes (EH-SN). The provisioning of harvesting capability (e.g. solar, wind or heat harvesting) and a rechargeable battery removes the fixed upper limit on the energy budget. Indeed, an EH-SN can constantly recharge during its lifetime. However, energy management is still required to balance energy harvesting and energy expenditure. Moreover, external conditions from which the harvesting circuitry draws its energy fluctuate over time. As a result, the amount of energy that can be harvested also varies largely with time. From a performance evaluation point of view, the harvesting process is an additional source of uncertainty which can affect performance.

This paper proposes a stochastic model to study the interplay between energy harvesting and SN performance. In particular, we adopt the hybrid WSN of [14], which consists of static sensors responsible for sensing environmental variables, and mobile sensors, referred to as IoT mobile sinks, that move to various designated sink locations where they gather the data that was sensed by static sensors. These mobile sinks were introduced to overcome the hot-

spot effect in sensor networks, see e.g. [15]. Both static and mobile sink nodes collect data from sensor nodes and sometimes act as gateways to other users by processing and sending relevant information. If all sensor data is relayed by the sensor nodes to a (static) sink node, nodes closer to the static sink are more heavily loaded as they need to relay more packets to the static sink in comparison with nodes further away. As a result, they consume more energy and may die at an early stage, or will frequently run out of energy if they can harvest energy. Mobile sinks overcome this problem by moving the sink around. See e.g. [16] for a discussion on design issues and challenges in existing distributed protocols for mobile sinks. Although mobility increases the network lifetime by balanced utilisation of power [17], it also introduces new challenges as connectivity to the mobile sinks is intermittent while the packet delivery delay should be sufficiently small [18].

Finally, we mention some applications of hybrid WSNs. Ren et al. optimise data collection scheduling of a hybrid WSN, and propose such a WSN for highway traffic surveillance and ecosystem monitoring [19]. In the highway traffic scenario, sensors are deployed along a highway and collect traffic information such as the number of vehicles and their speeds, the types of vehicles, etc. The ecosystem scenario includes monitoring of exotic plant growth or endangered animals. In a typical ecosystem monitoring scenario, humans or vehicles can only access the system via limited roads, while the exact sensor locations in the forest are not easily reached. In unmanned agriculture, Huang and Chang [20] propose a wireless sensor network with a mobile sink for collecting image data. The mobile sink mitigates the need to relay large amounts of data or to transmit large amounts over a large distance. In either case, considerable energy is required which cannot be delivered by the sensors. Similarly, Yang and Miao [21] propose a mobile sink to solve the problem of poor scalability and unbalanced energy consumption in farmland WSNs. Finally, motivated by energy constraints for the sensor nodes, Taherian et Al. [22] propose a WSN with a mobile sink for monitoring a railway transportation system.

1.1 Related literature

As the present paper investigates a Markovian model for an EH-WSN, we now discuss some related Markovian models. A first set of models consider the battery dynamics of a wireless sensor node, without data buffering. The battery state is usually discretised, meaning that the battery provides chunks of energy, rather than a continuous stream of energy. The simplest Markovian model of the battery, is a continuous or discrete-time (quasi-) birth-death Markov process, where births and deaths correspond to energy harvesting and energy expenditure respectively, see e.g. [23] where harvesting and recharging are combined and [24] where the harvesting process exhibits time correlation. The latter authors model the energy harvesting process as a two-state Markov modulated model, i.e., a node either harvests energy leading to a random increase of the amount of energy or is unable to harvest any energy. Also [25]

considers a two-state harvesting process and determines a transmission policy such that energy harvesting and consumption are balanced. While two-state Markov models already exhibit some time correlation, more realistic models for the time correlation in harvesting processes are studied in [26] and [27]. These authors study traces of solar harvesting processes and statistically verify that a Markov modulated process can be used to model the solar energy harvesting process. Relaxing the assumption that the sensor node remains in its different operational modes for an exponential amount of time, a semi-Markov model for the battery dynamics is studied in [28].

While the former models account for energy storage, data buffering is not considered. Data buffering however is accounted for in [29–33]. In [32], Gelenbe considers a model with Poisson arrivals of data and energy. Assuming that the time to transmit a packet is far smaller than the time to harvest the energy for the transmission, there is either no data or no energy at the node, which simplifies the analysis considerably. In [31], a similar model is studied where the sensor nodes are also subject to ongoing energy loss through standby power consumption and leakage from batteries and capacitors, and where transmissions can be corrupted by noise and mutual interference. If transmissions can be neither considered to happen almost immediately nor postponed due to unavailability of the receiver, both the data and energy queue can be simultaneously non-empty. Therefore, Markov models with two queues (an energy and a data queue) are required. In [29,30] such a Markov model is studied where energy and packet arrivals depend on an exogenous Markovian background process. This allows to include correlation in both harvesting, sensing and transmission processes. A similar Markov process is studied in [34] in the context of a “green” base station. Such a base station uses renewable energy sources for powering its operations. A somewhat different approach is proposed in [33]. In contrast, to the models above, these authors do not adopt energy chunks but model the battery as a fluid queue. The model allows for correlated energy harvesting (a fluid Markov process with two states), packet queueing and retransmissions, a sleep period and temporal death of the node. A temporal death state is reached when the node runs out of energy.

Various authors have also tackled control problems for EH-SNs by formulating the dynamics and control as a Markov decision process (MDP). A survey [35] reviews numerous applications of MDPs in EH-WSN and discusses and compares various algorithms and solution methods. We discuss some more recent contributions. In [36], Rao et Al. use the framework of Markov decision processes to determine the optimal task scheduling for a sensor node. Tasks include sensing, reading and writing from flash storage, packet reception and transmission, as well as computation. The MDP formulation accounts for task priorities and deadlines. Although energy harvesting is making WSNs self-sustainable, Lei et Al. [37] note that the uncertainty on harvesting leads to unreliability and instability which is becoming a major challenge in the design of networks. These authors formulate the problem as a constrained MDP and focus on scheduling algorithms that minimise data loss under delay constraints. Zordan et al. [38] investigate energy-aware lossy data compression

policies for SNs, and model the SN's transmission and energy dynamics as a constrained Markov decision problem. If a SN does not have sufficient energy to perform its tasks, a significant amount of energy can be saved by temporarily powering off. An MDP formulation for similar sleep and wake-up strategies is considered in [39]. Finally, accounting for both data- and energy buffers, an MDP formulation for deciding on transmissions for EH-WSNs with a mobile sink is proposed in [40].

1.2 Contributions

In contrast to the WSN models above, the present study focuses on controlling the collection frequency of the mobile sink in a hybrid WSN. Key to our study is the concept of "Quality of Information" which relates to the value the information brings to decision making. Combining sensor networks with data analytics enable fast data-to-decision applications that act in real time on the collected data such that the value the information brings to the decision not only depends on the quality and quantity but also on the timeliness of the information. Therefore, analogous to Quality of Service which measures the performance of a data communications network, the term "Quality of Information" (QoI) has been introduced to evaluate performance of sensor networks [41–43]. The present paper builds upon our previous paper [44] which considered a SN in a hybrid WSN, where energy harvesting and data collection does not exhibit time correlation. We summarise our contribution as follows:

- We propose a stochastic performance model for an energy harvesting sensor node in isolation. The performance of the sensor node depends on three exogenous processes: (i) the sensing process which describes the value of the information that is sensed; (ii) the energy harvesting process; and (iii) the data collection process which models the presence of the mobile sink such that the SN can offload its information.
- We introduce time correlation in the energy harvesting process to capture the time varying nature of harvesting source more realistically.
- We show that the model allows for studying the trade-off between the cost of frequent data collection — data collection is not only costly but also induces increased energy consumption at the SNs — and timely data delivery. The modelling assumptions allow for time correlation in the energy harvesting process, as well as for time correlation in the data sensing process.
- We present a numerical example which investigates the SN performance in terms of various system parameters including the polling probability of the mobile sink, the battery capacity, the parameters of the harvesting process, etc. In particular, the numerical section discusses how correlation in the harvesting process affects performance.

1.3 Organisation

The remainder of this paper is organised as follows. Sections 2 and 3 introduce the discrete-time stochastic model and its analysis, respectively. In section 4, we focus on optimal data collection at EH-SNs with time-correlated energy harvesting by means of some numerical examples. Finally, we conclude and summarise our results in section 5.

2 Mathematical model

We consider an energy harvesting sensor node. The node is equipped with a battery for storing harvested chunks of energy and on-board memory to store sensed information. Time is assumed to be discrete, i.e., time is divided into fixed length intervals or slots and we study the evolution of the value of information of the sensed data and the energy levels at slot boundaries. During each slot, sensor data is collected and stored in on-board memory, and energy is collected and stored in the battery. Whenever the mobile sink passes by, the data (if any) is transferred to the mobile sink, provided the sensor node has sufficient energy to transmit the data and there is any data to transmit. Hence, three interacting random processes — the energy harvesting process, the sensor data arrival process and the data collection process — determine the evolution of the state of the sensor node over time. These are described underneath.

2.1 Energy harvesting process

Adopting the energy chunk paradigm, the battery can store at most C chunks of energy, and the amount of energy that is harvested during a slot is a discrete random variable. We assume that M energy chunks are required to sense data and N chunks are required for a transmission to the mobile sink. Sensing has priority over transmissions, meaning that the node senses when there are M chunks present and transmits when there are $M + N$ energy chunks present when the mobile sink is in range.

We consider a Markov-modulated energy harvesting process, where the modulating process takes values in the set $\mathcal{J} = \{1, \dots, J\}$. Let $E_n \in \mathcal{J}$ denote the state of the modulating process during slot n , and let H_n denote the number of chunks of energy that are harvested in this slot. The energy harvesting process is then characterised by the following marked state transition probabilities,

$$\alpha_{ij}(m) = \mathbb{P}[H_n = m, E_{n+1} = j | E_n = i].$$

That is, $\alpha_{ij}(m)$ is the transition probability to go from state i to state j while harvesting m chunks of energy. For further use, we also define the transition

probabilities $\bar{\alpha}_{ij}^{(m)}$ from state i to state j while harvesting at least m chunks,

$$\bar{\alpha}_{ij}(m) = \mathbb{P}[H_n \geq m, E_{n+1} = j | E_n = i] = \sum_{\ell=m}^{\infty} \alpha_{ij}(\ell).$$

2.2 Sensor data arrival process

The sensor node is equipped with on-board memory to temporarily store the information sensed during the consecutive time slots. Assuming that sensing data does not require much storage, we do not track the size of the sensor data, but track the value (or the quality) of information of the data instead. Note that in some specific cases, the size of the data can even be constant over time. This is e.g. the case if the data is a vector of environmental quantities which is regularly updated by the sensing process. The amount of information that is described by this vector however depends on the actual data values and can differ considerably over time.

Let S_n denote the value of the data sensed during the n th slot. We assume that the mean sensed value of information is constant over time i.e., $\bar{S} = \mathbb{E}[S_n]$ for $n \in \mathbb{N}$, as well as the following independence assumption: The indicator that there is no sensed data $\mathbf{1}_{\{S_n=0\}}$ constitutes a sequence of independent and identically distributed random variables with mean s_0 , s_0 being the probability that there is no information. Note that the latter assumption does not exclude correlation in the sensing process. Indeed, let $\{\hat{S}_n\}$ be a sequence of stationary ergodic random variables with mean $\bar{S}/(1-s_0)$, and let $\{I_n\}$ be a sequence of Bernoulli random variables with mean s_0 , then $\{\hat{S}_n(1-I_n)\}$ is a sequence of non-independent random variables satisfying the independence assumption above.

Assuming that older data is less relevant to decision making, the value of the information drops while it is not collected. To capture such loss, we assume that the value of information at the sensor node is discounted in each time slot with discount factor γ . Moreover, we assume that the value of information is additive: the value of the data sensed during a time slot that cannot be transmitted is added to the (discounted) value already at the sensor.

2.3 Data collection process

Finally we assume that the time (in slots) between data collection constitutes a sequence of independent and identically geometrically distributed random variables with success probability p . Let P_n be the binary random variable which denotes whether the data is collected at the n th slot boundary or not. Then with the assumption above, the process $\{P_n\}$ constitutes a Bernoulli process with $\mathbb{P}[P_n = 1] = \mathbb{E}[P_n] = p$.

The data collection process models the availability of the sink at the location of the sensor node and therefore relates to the trajectory that is followed

by the mobile sink. I.e., $P_n = 1$ if the sink is in the transmission range in slot n while $P_n = 0$ if this is not the case. The probability p therefore corresponds to the fraction of time that the mobile sink is in range during the trajectory of the mobile sink.

3 Performance analysis

We now study the evolution of the battery level and the value of information. We first describe the stochastic difference equations for the battery level and the value of information, and then calculate the battery level probabilities and the mean value of information.

3.1 Stochastic difference equations

Let B_n denote the battery level at the beginning of time slot n and let V_n denote the value of information at the sensor node at the n th slot boundary.

Assuming that the energy harvested in slot n cannot be used for sensing and transmitting data in slot n , and that any energy which cannot be stored in the battery is lost, we have,

$$B_{n+1} = \min(B_n - M1_{\{B_n \geq M\}} - NT_n + H_n, C).$$

Here T_n is the binary random variable which indicates if there is a transmission during slot n or not. We express T_n in terms of the battery level B_n , the value of information V_n , and the indicator of the mobile sink P_n below.

As the value of information is discounted with discount factor γ and information is additive, we have,

$$V_{n+1} = \gamma V_n(1 - T_n) + S_n 1_{\{B_n \geq M\}}.$$

Here we assumed that all information is immediately transferred to the mobile sink when there is a transmission ($T_n = 1$), while any newly sensed data S_n is not yet available for transmission.

When data is collected, the sensor node always transmits whenever there is data and energy, independently of the value of information. This is a natural assumption when the value of information cannot easily be assessed at the sensor node. For example, the value may only become apparent when the information of multiple sensor nodes is combined. To facilitate analysis, let A_n be the binary random variable which indicates whether there is information at the sensor node or not, that is, $A_n = 1_{\{V_n > 0\}}$. Here, the indicator function $1_{\{\cdot\}}$ evaluates to 1 if its argument is true and to 0 if this is not the case. While A_n can be directly expressed in terms of V_n , the following recursion for A_n simplifies the analysis,

$$A_{n+1} = 1_{\{S_n > 0\}} 1_{\{B_n \geq M\}} + (1 - 1_{\{S_n > 0\}} 1_{\{B_n \geq M\}}) A_n (1 - T_n).$$

The first and second term correspond to the case that information was and was not sensed during the preceding slot, respectively. In the latter case, there is only information at the sensor node if there was already information at the preceding slot boundary that was not transmitted. Notice that the evolution of A_n only depends on the indicator that there is information $1_{\{S_n > 0\}}$, and not on the actual value of S_n .

Finally, we express T_n in terms of A_n , B_n and P_n as follows,

$$T_n = A_n 1_{\{B_n \geq M+N\}} P_n.$$

That is, there is a transmission provided there is data to transmit ($A_n = 1$), energy to transmit ($B_n \geq M + N$), and an opportunity to transmit ($P_n = 1$).

3.2 Battery level probabilities

We first focus on expressions for the battery level probabilities. To this end, we note that the process $\{(B_n, A_n, E_n), n \in \mathbb{N}\}$ is a Markov process, and we study its stationary distribution. Let

$$\tilde{b}_j(k) = \lim_{n \rightarrow \infty} \mathbb{P}[B_n = k, A_n = 0, E_n = j]$$

be the stationary probability that there are k chunks of energy in the battery, while there is no information at the sensor node ($A_n = 0$), and the energy harvesting process is in state $E_n = j$, $j \in \mathcal{J}$. The stationary battery level probabilities

$$\hat{b}_j(k) = \lim_{n \rightarrow \infty} \mathbb{P}[B_n = k, A_n = 1, E_n = j]$$

when there is information at the sensor node are defined likewise. Note that we account for the availability of information as the decision to transmit depends on the availability of information, which in turn affects the battery level.

Expressing the state probabilities at slot $n + 1$ in terms of those at slot n by conditioning on the values of A_n , P_n , E_n and H_n , and then sending n to ∞ yields the following set of balance equations,

$$\begin{aligned} \tilde{b}_j(k) = & \sum_{i=1}^J \sum_{\ell=0}^{M-1} \tilde{b}_i(\ell) \alpha_{ij}(k, \ell) + s_0 \sum_{i=1}^J \sum_{\ell=M}^C \tilde{b}_i(\ell) \alpha_{ij}(k, \ell - M) \\ & + p s_0 \sum_{i=1}^J \sum_{\ell=M+N}^C \hat{b}_i(\ell) \alpha_{ij}(k, \ell - M - N), \end{aligned} \quad (1)$$

and,

$$\begin{aligned} \hat{b}_j(k) = & \sum_{i=1}^J \sum_{\ell=0}^{M-1} \hat{b}_i(\ell) \alpha_{ij}(k, \ell) + \sum_{i=1}^J \sum_{\ell=M}^C \left(\hat{b}_i(\ell) + (1 - s_0) \tilde{b}_i(\ell) \right) \alpha_{ij}(k, \ell - M) \\ & + p \sum_{i=1}^J \sum_{\ell=M+N}^C \hat{b}_i(\ell) ((1 - s_0) \alpha_{ij}(k, \ell - M - N) - \alpha_{ij}(k, \ell - M)), \end{aligned} \quad (2)$$

for $k = 0, 1, \dots, C$ with $j \in \mathcal{J}$. Here, we introduced for $i, j \in \mathcal{J}$ and $\ell \leq k$,

$$\alpha_{ij}(k, \ell) = \begin{cases} \tilde{\alpha}_{ij}(k - \ell) & \text{for } k = C, \\ \alpha_{ij}(k - \ell) & \text{otherwise.} \end{cases}$$

We collect the probabilities $\tilde{b}_j(k)$ and $\hat{b}_j(k)$ in the row vectors $\tilde{\mathbf{b}}(k) = [\tilde{b}_j(k)]_{j=1}^J$ and $\hat{\mathbf{b}}(k) = [\hat{b}_j(k)]_{j=1}^J$ for each k , and introduce the row vector $\mathbf{b}(k) = [\tilde{\mathbf{b}}(k), \hat{\mathbf{b}}(k)]$. The equations above can then be summarised in the following matrix equation,

$$\mathbf{b}(k) = \sum_{\ell=0}^C \mathbf{b}(\ell) \mathcal{B}(\ell, k) \quad (3)$$

where $\mathcal{B}(\ell, k) = [\mathcal{B}_{ij}(\ell, k)]_{i,j=1}^2$ with,

$$\begin{aligned} \mathcal{B}_{11}(\ell, k) &= \mathbf{1}_{\{\ell < M\}} \mathcal{A}(k, \ell) + s_0 \mathbf{1}_{\{\ell \geq M\}} \mathcal{A}(k, \ell - M), \\ \mathcal{B}_{21}(\ell, k) &= ps_0 \mathbf{1}_{\{\ell \geq M+N\}} \mathcal{A}(k, \ell - M - N), \\ \mathcal{B}_{12}(\ell, k) &= \mathbf{1}_{\{\ell \geq M\}} (1 - s_0) \mathcal{A}(k, \ell - M), \\ \mathcal{B}_{22}(\ell, k) &= \mathbf{1}_{\{\ell < M\}} \mathcal{A}(k, \ell) + \mathbf{1}_{\{\ell \geq M\}} \mathcal{A}(k, \ell - M) \\ &\quad + p \mathbf{1}_{\{\ell \geq M+N\}} ((1 - s_0) \mathcal{A}(k, \ell - M - N) - \mathcal{A}(k, \ell - M)). \end{aligned}$$

In the former expressions $\mathcal{A}(k, \ell)$ is the $J \times J$ matrix with elements $\alpha_{ij}(k, \ell)$, $i, j \in \mathcal{J}$. Finally, let $\mathbf{b} = [\mathbf{b}(k)]_{k=0}^C$ and $\mathcal{B} = [\mathcal{B}(\ell, k)]_{\ell, k=0}^C$ such that (3) immediately yields $\mathbf{b} = \mathbf{b} \mathcal{B}$. The stationary probability vector \mathbf{b} is the normalised solution of this matrix equation.

It is easy to check that $\mathcal{B}(\ell, k)$ is a zero matrix for $k < \ell - M - N$. Indeed, the battery level drops for at most $M + N$ levels during a slot. Hence, \mathcal{B} has an upper-Hessenberg block structure with blocks of size $(2J(M + N)) \times (2J(M + N))$. This block structure can then be exploited to calculate the stationary probability vector \mathbf{b} , for example by using linear level reduction (see e.g. [45]).

3.3 Mean value of information

We now study the mean value of information at the sensor node. Noting that $A_n = 0$ implies $V_n = 0$, we obviously have,

$$\mathbb{E}[V_n \mathbf{1}_{\{B_n=k, A_n=0, E_n=j\}}] = 0,$$

for $k = 0, 1, \dots, C$ and $j \in \mathcal{J}$. Let,

$$v_j(k) \doteq \lim_{n \rightarrow \infty} \mathbb{E}[V_n \mathbf{1}_{\{B_n=k, A_n=1, E_n=j\}}]$$

be the mean value information when there is information, when there are k chunks of energy and the harvesting process is in state j , $k = 0, 1, \dots, C$ and $j \in \mathcal{J}$. Furthermore, let $v(k) = \sum_{j=1}^J v_j(k)$ be the mean value information when there is information and there are k chunks of energy.

As for the battery level probabilities, we again condition on the values of B_n , A_n and P_n , to express the mean $\mathbb{E}[V_{n+1} \mathbf{1}_{\{B_{n+1}=k, A_{n+1}=1, E_{n+1}=j\}}]$ in terms of the mean $\mathbb{E}[V_n \mathbf{1}_{\{B_n=k, A_n=1, E_n=j\}}]$ and the battery level probabilities. After taking the limit $n \rightarrow \infty$, we find

$$\begin{aligned} v_j(k) = & \sum_{i=1}^J \sum_{\ell=M}^C (\gamma v_i(\ell) + \bar{S}(\hat{b}_i(\ell) + \tilde{b}_i(\ell))) \alpha_{ij}(k, \ell - M) \\ & + \gamma \sum_{i=1}^J \sum_{\ell=0}^{M-1} v_i(\ell) \alpha_{ij}(k, \ell) + p \bar{S} \sum_{i=1}^J \sum_{\ell=M+N}^C \hat{b}_i(\ell) \alpha_{ij}(k, \ell - M - N) \\ & - p \sum_{i=1}^J \sum_{\ell=M+N}^C (\gamma v_i(\ell) + \bar{S} \hat{b}_i(\ell)) \alpha_{ij}(k, \ell - M), \quad (4) \end{aligned}$$

for $k \in \{0, 1, \dots, C\}$. Collecting the $v_j(k)$ for $j \in \mathcal{J}$ in the row vector $\mathbf{v}(k) = [v_j(k)]_{j=1}^J$, we obtain the system of vector equations,

$$\mathbf{v}(k) = \gamma \sum_{\ell=0}^C \mathbf{v}(\ell) \mathcal{V}(\ell, k) + \bar{S} \mathbf{w}(k), \quad (5)$$

for $k \in \{0, 1, \dots, C\}$ with

$$\mathcal{V}(\ell, k) = \mathbf{1}_{\{\ell < M\}} \mathcal{A}(k, \ell) + \mathbf{1}_{\{\ell \geq M\}} \mathcal{A}(k, \ell - M) - p \mathbf{1}_{\{\ell \geq M+N\}} \mathcal{A}(k, \ell - M)$$

and

$$\begin{aligned} \mathbf{w}(k) = & \sum_{\ell=M}^C (\hat{\mathbf{b}}(\ell) + \tilde{\mathbf{b}}(\ell)) \mathcal{A}(k, \ell - M) \\ & + p \sum_{\ell=M+N}^C \hat{\mathbf{b}}(\ell) (\mathcal{A}(k, \ell - M - N) - \mathcal{A}(k, \ell - M)) \end{aligned}$$

As for the battery probabilities, the equations for the mean value of information is a set of $J(C+1)$ equations with as many unknowns and is easily solved provided C is not prohibitively large. For larger C , one can again exploit the upper-Hessenberg block structure of the set of equations, the blocks now being $J(M+N) \times J(M+N)$ matrices.

3.4 Cost and profit of collection

Once the mean value of information at the sensor node for the different battery levels is determined by numerically solving the systems of equations in the preceding subsections, we can determine the cost or profit of data collection. Noting that data is only collected if there is sufficient energy and the mobile sink passes by, we find that the profit of the data collection, i.e. the mean value

of information of the sensor data that is actually collected (per time slot), can be expressed in terms of the v_k 's as follows,

$$\bar{V} = p \sum_{k=M+N}^C \sum_{j=1}^J v_j(k).$$

Assuming that there is a fixed cost for every time the mobile sink collects data from the sensor, the mean cost of collection is linear in p

$$\bar{C} = cp$$

where c is a constant cost, expressed in terms of the value of information. Subtracting the cost of collection from the mean value of information that is collected, we find that profit or the mean value of the data collection equals

$$\bar{V}_p = -cp + \bar{V}.$$

Remark 1 Recall that the collection cost relates to the trajectory that is taken by the mobile sink: p is the probability that the sink is available, or the fraction of time that the sink is in range. If the mobile sink moves continuously, a higher p means that the mobile sink will visit the sensor node more often, while if the mobile sink is not constantly moving, increasing p means that the mobile sink remains idle for less time. In either case there is a cost increase, which we assume to be proportional to p .

Remark 2 While we mainly focus on a single node, it is worth pointing out that the results for the single sensor node can also be used to find the optimal trajectory of the mobile sink. By carefully analysing a given trajectory of a mobile sink that covers a set of sensor nodes, one can find the fraction of time the mobile sink is in range for every sensor node. Say p_k is this fraction for the k th sensor node. Given p_k , the calculations above allow for finding the value of information at the k th sensor node. It then remains to find the trajectory of the mobile sink which maximises the sum of the values of information.

4 Numerical results and discussion

Having established a numerical procedure to calculate the mean value of the data collection in the preceding section, we now illustrate our approach by some numerical examples. In particular, we study the mean value of the data collection \bar{V}_p in terms of the collection probability p as well as the optimal collection probability in terms of various system parameters, including the effect of time correlation in the energy harvesting process. Note that the optimal collection probability below is only locally optimal. That is, it is the collection probability which optimises the value of information of the single sensor node under consideration, while accounting for some collection cost.

For the numerical examples, we assume that the harvesting process is modulated by a Markov process with two states, $\mathcal{J} = \{1, 2\}$. The harvesting process is inactive in state 1 while the number of energy chunks constitutes a sequence of independent Poisson distributed random variables with mean λ when the modulating process is in state 2. Let α be the transition probability from the active to the inactive state and let β be the transition probability from the inactive to the active state. Moreover, let $h_m = \exp(-\lambda)\lambda^m/m!$. Then we have $\alpha_{11}(m) = (1 - \beta)1_{\{m=0\}}$, $\alpha_{12}(m) = \beta 1_{\{m=0\}}$, $\alpha_{21}(m) = \alpha h_m$ and $\alpha_{22}(m) = (1 - \alpha)h_m$.

The following alternative characterisation of the harvesting process will be used in the remainder. Let σ be the fraction of time the harvesting process is in the active state and let κ be the average duration of an active and an inactive period. The latter is a measure for the time-scale of the energy harvesting process, larger κ meaning that the harvesting process alternates more slowly between states. Simple calculations yield,

$$\sigma = \frac{\beta}{\alpha + \beta}, \quad \kappa = \frac{1}{\alpha} + \frac{1}{\beta}, \quad \lambda^* = \lambda \sigma = \frac{\lambda \beta}{\alpha + \beta}$$

For a fixed σ , any choice $\kappa \geq \max(\sigma^{-1}, (1 - \sigma)^{-1})$ uniquely defines α and β . In particular, for $\kappa = \sigma^{-1} + (1 - \sigma)^{-1}$ we have no time correlation as $\alpha = 1 - \beta$ such that the next state is active or inactive, independently from the current state.

We now define the parameter values used throughout this section. On average $\lambda = 5$ energy chunks are harvested in active slots, unless indicated otherwise. It takes $M = 1$ energy chunk to sense the data and $N = 4$ energy chunks for transmitting the information. The average value of the sensing data is $\bar{S} = 2$, while we assume that during most time slots there is no data: $s_0 = 0.9$. Moreover, we set the information decay rate equal to $\gamma = 0.9$ and the maximum capacity of the battery to $C = 16$, unless otherwise specified. Finally, the cost to collect the data is $c = 1$ for all plots.

Fig. 1 depicts the mean value of information \bar{V}_p in terms of the collection probability p . Energy harvesting is slightly bursty: we assume $\sigma = 0.4$ and $\kappa = 16$. Fig. 1(a) shows the mean value of information for different battery capacities C as indicated, while the decay rate is fixed to $\gamma = 0.9$. On the other hand, fig. 1(b) fixes the battery capacity to $C = 16$ and shows \bar{V}_p for different decay rates as indicated. From both figures, we see that $\bar{V}_p = \bar{V} - \bar{C}$ is a concave function of p . This observation can be explained by noting that an increase of p is beneficial for small p , as the data is collected more often (the increase of \bar{V} compensates the additional collection cost \bar{C}). For high p , it is however quite likely that there is not sufficient energy to transmit. Therefore, the possible gain of frequent data collection cannot compensate the collection cost (the increase of \bar{V} no longer compensates the additional collection cost \bar{C}). Further, fig. 1(a) shows that it is beneficial to increase the battery size. Having a battery with more capacity facilitates compensating periods with limited energy. Similarly, fig. 1(b) indicates that increasing the discounting factor also leads to better performance. A higher discounting factor implies that the value

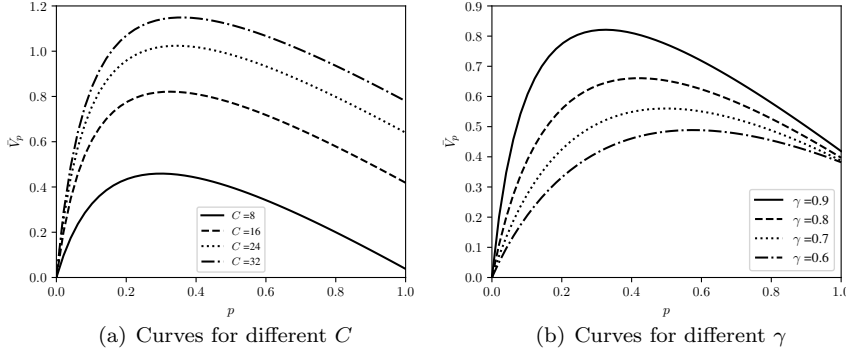


Fig. 1 The mean value of the data collection \bar{V}_p versus the polling probability p for (a) different values of the battery capacity C and (b) different values of the decay rate γ as indicated.

of information decays more slowly such that more information is available for collection. Of course, the discounting factor models the timeliness of the information and cannot be modified freely in practice. One may expect that for $p = 1$, \bar{V}_p does not depend on the discounting factor as data is constantly collected. This is however not the case, as a lack of energy prevents constant transmissions. Summarising, in selecting the collection probability p , there is a clear trade-off between the cost of frequent collection \bar{C} , and the collected information \bar{V} , while more battery capacity and higher discounting factors both lead to a higher mean value of information \bar{V}_p .

We now focus on the effect of varying the length of the active and inactive harvesting periods. Figure 2(a) depicts \bar{V}_p versus the polling probability p for different κ as indicated. Here, we fix the fraction of time the system is active: $\sigma = 0.4$. It is observed that increasing the length of the active and inactive periods affects \bar{V}_p adversely. For high values of κ , the system has longer active period followed by longer inactive periods. During long inactive periods, the battery drains completely as there is no energy harvesting while excess energy during active periods is lost as the battery capacity is limited. As a result, there will be less value of information to be collected by the mobile sink. Fig. 2(b) shows the mean value of the data collection versus the energy harvesting rate for different values of σ . We fix the data collection probability to $p = 0.2$ and the time-scale parameter to $\kappa = 16$. It can be clearly seen from the figure that the increase in harvesting capability improves the system performance. As κ is fixed, increasing σ means the system has longer active periods during which the battery level increases on average, and shorter inactive period for which the battery level decreases on average. Therefore, the chance of having sufficient energy for data transmissions increases, and therefore also the mean value of information collected by the mobile sink. However, if there is already sufficient energy for data sensing and transmissions, no further improvement is seen. In this case, the fixed battery capacity plays an important role. There

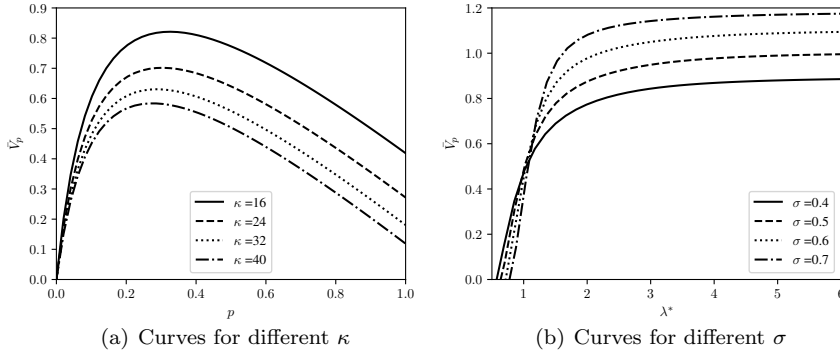


Fig. 2 The mean value of the data collection \bar{V}_p versus the polling probability for different energy harvesting process (a) and mean value of the data collection \bar{V}_p versus energy harvesting rate λ for different fractions of active time σ as indicated (b).

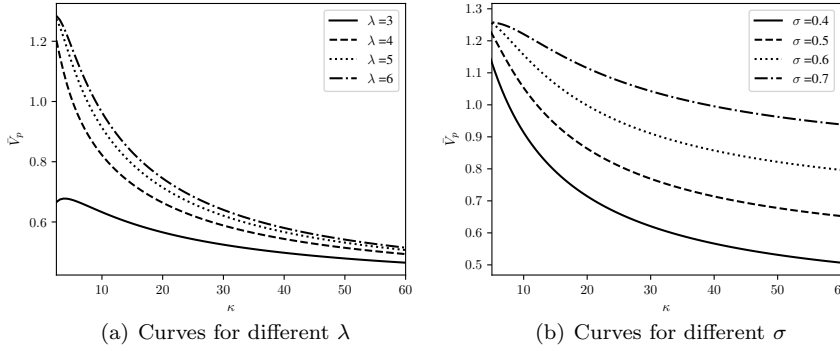


Fig. 3 The mean value of data collection \bar{V}_p versus average duration of active and inactive process κ for (a) different values of the energy harvesting rate λ and (b) different fractions of active time σ as indicated.

is no additional benefit from longer active periods for high λ , since the battery reaches its maximum capacity quickly and excess energy chunks are lost.

To better understand the influence of the time-scale parameter κ , Fig. 3 shows \bar{V}_p versus κ , for fixed $\sigma = 0.4$ and various λ as indicated (Fig. 3(a)) and for fixed $\lambda = 5$ and various σ as indicated (Fig. 3(b)). As in the preceding plot, the collection probability is fixed to $p = 0.2$. Fig. 3 confirms the observation that an increase of κ leads to a decrease of \bar{V}_p . Even when the harvesting capability is considerable, a performance degradation is unavoidable when harvesting is interrupted for a long time. This effect is mitigated if one increases the battery capacity, as a fully charged battery with higher capacity can sustain the operation of the SN for longer times without harvesting.

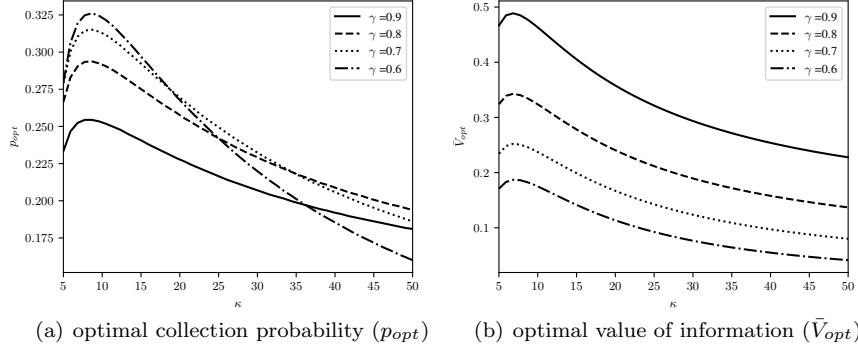


Fig. 4 The optimal collection probability p_{opt} (a) and the corresponding mean value of the collected information \bar{V}_{opt} (b) vs. the time-scale parameter κ for different γ as indicated.

The time-scale parameter κ also affects the value of the optimal collection probability p_{opt} . Fig. 4 depicts the value of the optimal collection probability p_{opt} (Fig. 4(a)) and the corresponding mean value of information \bar{V}_{opt} (Fig. 4(b)) versus κ for different values of the discounting factor γ as indicated. We set $\sigma = 0.2$ and $\lambda = 5$. For increasing κ , it is optimal to collect more if κ is small. In contrast, for higher κ -values, it is optimal to collect less if κ further increases. While this is not intuitively clear, simple arguments show that the optimal collection probability increases as long as the average information at the sensor node exceeds the collection cost. As the average information at the sensor decreases for increasing κ , the optimal collection probability first increases and then decreases. Similar arguments explain the difference in decay rate of the optimal collection probability for different discount factors γ . The resulting optimal mean value is easier explained: discounting less means that the information is longer available at the sensor node, and that more information is collected.

Finally, we investigate how the absence of information probability s_0 affects the optimal data collection probability and the corresponding value of information. In fig. 5, we vary s_0 while \bar{S} is fixed. Higher s_0 means there is less chance that information arrives in a slot. However, if there is data, the information it carries is more significant. In other words, the process S_n is more bursty. We again plot the optimal collection probability p_{opt} (Figure 5(a)) and the corresponding value of information \bar{V}_{opt} (Figure 5(b)). The harvesting process is characterised by $\sigma = 0.2$ and $\kappa = 16$. It is observed from Figure 5(a) that the optimal collection probability increases as the information process becomes bursty. Indeed, for increasing s_0 , the data should be collected more frequently as one does not want to miss out on the rare occasions that there is information.

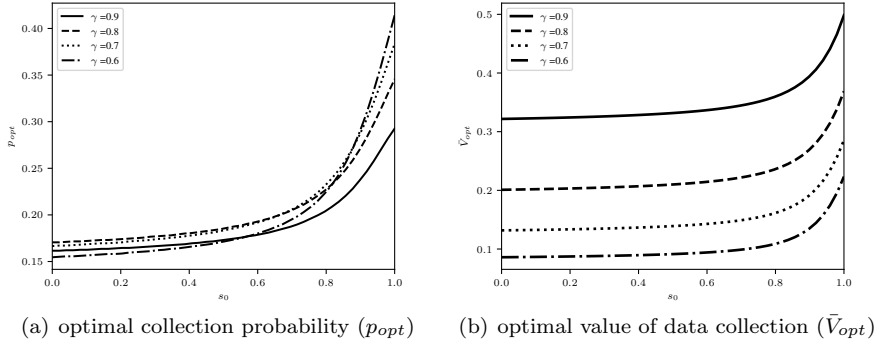


Fig. 5 The optimal collection probability p_{opt} (a) and the corresponding mean value of the collected information \bar{V}_{opt} (b) vs. the absence of information probability s_0 for different γ as indicated.

5 Conclusion

In this paper, we investigated the optimal data collection rate for an energy harvesting sensor node. The assumptions at hand allow for correlation in both the harvesting and sensing processes. The former is modelled by a Markovian arrival process. For the latter, we assume that the process is first-order stationary with an additional independence assumption for the related indicator that there is no information. We show that the mean value of the information that is collected can be found by solving a system of linear equations. We find the upper-Hessenberg block-matrix representation of this set of equations, which can be exploited to speed up the calculations. To illustrate our approach, we have presented a number of numerical examples, which exemplify how the various system parameters affect performance. In particular, it is demonstrated that the presence of time correlation in the harvesting process has a major impact on performance.

References

1. L. Atzori, A. Iera, and G. Morabito. The internet of things: a survey. *Computer Networks*, 54(15):2787–2805, 2010.
2. J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami. Internet of things (IoT): a vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29:1645–1660, 2013.
3. I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38(4):393–422, 2002.
4. I. F. Akyildiz and M. C. Vuran. *Wireless sensor networks*. Wiley, 2010.
5. H. M. Ammari, N. Gomes, M. Jacques, B. Maxim, and D. Yoon. A survey of sensor network applications and architectural components. *Ad Hoc & Sensor Wireless Networks*, 25(1–2):1–44, 2015.
6. H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelfflé. *Vision and challenges for realising the internet of things*. European commission, 2010.

7. S. Sudevalayam and P. Kulkarni. Energy harvesting sensor nodes: Survey and implications. *IEEE Communications Surveys & Tutorials*, 13(3):443–461, 2011.
8. J. Polastre, J. Hill, and D. Culler. Versatile low power media access for wireless sensor networks categories and subject descriptors. In *Proc. of the 2nd International Conference on Embedded Networked Sensor Systems (SenSys '04)*, pages 95–107, 2004.
9. W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. In *Proc. of the 33rd Annual Hawaii International Conference on System Sciences*, pages 3005–3014, 2000.
10. G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella. Energy conservation in wireless sensor networks: A survey. *Ad Hoc Networks*, 7(3):537–568, 2009.
11. S. Ganeriwal, I. Tsigkogiannis, H. Shim, V. Tsiatsis, M. B. Srivastava, and D. Ganesan. Estimating clock uncertainty for efficient duty-cycling in sensor networks. *IEEE/ACM Transactions on Networking*, 17(3):843–856, 2009.
12. X. Wang, G. Xing, Y. Zhang, C. Lu, R. Pless, and C. Gill. Integrated coverage and connectivity configuration in wireless sensor networks. In *Proc. of the First International Conference on Embedded Networked Sensor Systems (SenSys '03)*, pages 28–39, 2003.
13. J. A. Paradiso and T. Starner. Energy scavenging for mobile and wireless electronics. *IEEE Pervasive Computing*, 4(1):18–27, 2005.
14. Z. Zhou, C. Du, L. Shu, G. Hancke, J. Niu, and H. Ning. An energy-balanced heuristic for mobile sink scheduling in hybrid WSNs. *IEEE Transactions on Industrial Informatics*, 12(1):28–40, 2016.
15. Y. Bi, J. Niu, L. Sun, W. Huangfu, and Y. Sun. Moving schemes for mobile sinks in wireless sensor networks. In *Proc. of the IEEE International Performance Computing, and Communications Conference*, pages 101–108, 2007.
16. C. Tunca, S. Isik, M. Y. Donmez, and C. Ersoy. Distributed mobile sink routing for wireless sensor networks: A survey. *IEEE Communications Surveys Tutorials*, 16(2):877–897, 2014.
17. B. Nazir and H. Hasbullah. Mobile sink based routing protocol (MSRP) for prolonging network lifetime in clustered wireless sensor network. In *Proc. of the 2010 International Conference on Computer Applications and Industrial Electronics*, pages 624–629, 2010.
18. Y. Gu, Y. Ji, J. Li, and B. Zhao. ESWC: Efficient scheduling for the mobile sink in wireless sensor networks with delay constraint. *IEEE Transactions on Parallel and Distributed Systems*, 24(7):1310–1320, 2013.
19. X. Ren, W. Liang, and W. Xu. Data collection maximization in renewable sensor networks via time-slot scheduling. *IEEE Transactions on Computers*, 64(7):1870–1883, 2015.
20. S.-C. Huang and H.-Y. Chang. A farmland multimedia data collection method using mobile sink for wireless sensor networks. *MULTIMEDIA TOOLS AND APPLICATIONS*, 76(19):19463–19478, 2017.
21. Y. Yang and Y. Miao. A path planning method for mobile sink in farmland wireless sensor network. In *Proc. of the IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference*, pages 1157–1160, 2017.
22. M. Taherian, M. Maen, and M. Haghparast. Promoting the quality level of signaling in railway transportation system taking advantage from wireless sensor networks technology. *COMPUTERS*, 6(3), 2017.
23. J. Lei, R. Yates, and L. Greenstein. A generic model for optimizing single-hop transmission policy of replenishable sensors. *IEEE Transactions on Wireless Communications*, 8(2):547–551, 2009.
24. A. Seyedi and B. Sikdar. Modeling and analysis of energy harvesting nodes in wireless sensor networks. In *Proc. of the Forty-Sixth Annual Allerton Conference*, Allerton House, IL, USA, 2008.
25. N. Michelusi, K. Stamatiou, and M. Zorzi. Transmission policies for energy harvesting sensors with time-correlated energy supply. *IEEE Transactions on Communications*, 61(7):2988–3001, 2013.
26. P. Lee, Z. A. Eu, M. Han, and H. P. Tan. Empirical modeling of a solar-powered energy harvesting wireless sensor node for time-slotted operation. In *2011 IEEE Wireless Communications and Networking Conference*, pages 179–184, 2011.

27. C. K. Ho, P. D. Khoa, and P. C. Ming. Markovian models for harvested energy in wireless communications. In *Proc. of the 2010 IEEE International Conference on Communication Systems (ICCS)*, pages 311–315, Nov 2010.
28. M. Y. Naderi, S. Basagni, and K. R. Chowdhury. Modeling the residual energy and lifetime of energy harvesting sensor nodes. In *Proc. of the 2012 IEEE Global Communications Conference (GLOBECOM)*, 2012.
29. E. De Cuyper, K. De Turck, and D. Fiems. Stochastic modelling of energy harvesting for low power sensor nodes. In *Proc. of the 7th Int. Conference on Queueing Theory and Networking Applications (QTNA 2012)*, 2012.
30. E. De Cuyper, K. De Turck, and D. Fiems. A queueing model of an energy harvesting sensor node with data buffering. *Telecommunication Systems*, 67(2):281–295, 2018.
31. E. Gelenbe and Y. M. Kadioglu. Energy loss through standby and leakage in energy harvesting wireless sensors. In *2015 IEEE 20th International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD)*, pages 231–236, Sept 2015.
32. E. Gelenbe. A sensor node with energy harvesting. *SIGMETRICS Performance Evaluation Review*, 42(2):37–39, 2014.
33. L. Tan and S. Tang. Energy harvesting wireless sensor node with temporal death: Novel models and analyses. *IEEE/ACM Transactions on Networking*, 25(2):896–909, 2017.
34. I. Dimitriou, S. Alouf, and A. Jean-Marie. A Markovian queueing system for modeling a smart green base station. In *Proc of the 12th European Workshop on Computer Performance Engineering (EPEW 2015)*, pages 3–18, Madrid, Spain, 2015.
35. M. Abu Alsheikh, D. T. Hoang, D. Niyato, H. P. Tan, and S. Lin. Markov decision processes with applications in wireless sensor networks: A survey. *IEEE Communications Surveys Tutorials*, 17(3):1239–1267, 2015.
36. V. S. Rao, R. V. Prasad, and I. G. M. M. Niemegeers. Optimal task scheduling policy in energy harvesting wireless sensor networks. In *Proc. of the 2015 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1030–1035, 2015.
37. L. Lei, Y. Kuang, X. S. Shen, K. Yang, J. Qiao, and Z. Zhong. Optimal reliability in energy harvesting industrial wireless sensor networks. *IEEE Transactions on Wireless Communications*, 15(8):5399–5413, Aug 2016.
38. D. Zordan, T. Melodia, and M. Rossi. On the design of temporal compression strategies for energy harvesting sensor networks. *IEEE Transactions on Wireless Communications*, 15(2):1336–1352, Feb 2016.
39. M. Mitici, J. Goseling, M. de Graaf, and R. J. Boucherie. Energy-efficient data collection in wireless sensor networks with time constraints. *Performance Evaluation*, 102:34–52, 2016.
40. K. Patil, K. De Turck, and D. Fiems. A two-queue model for optimising the value of information in energy-harvesting sensor networks. *Performance Evaluation*, 119:27–42, 2018.
41. V. Sachidananda, A. Khelil, and N. Suri. Quality of information in wireless sensor networks: A survey. In *Proc. of the 15th International Conference on Information Quality (ICIQ'10)*, pages 193–207, 2010.
42. C. Bisdikian, L.M. Kaplan, and M.B. Srivastava. On the quality and value of information in sensor networks. *ACM Transactions on Sensor Networks*, 9(4):48, 2013.
43. E.C.-H. Ngai and P. Gunningberg. Quality-of-information-aware data collection for mobile sensor networks. *Pervasive and Mobile Computing*, 11:203–215, 2014.
44. K. Patil, K. De Turck, and D. Fiems. Optimal data collection in hybrid energy-harvesting sensor networks. In *Proc. of the 23rd International Conference on Analytical and Stochastic Modelling Techniques and Applications (ASMTA2016)*, Cardiff, Wales, 2016.
45. C. Blondia and O. Casals. Statistical multiplexing of VBR sources: A matrix-analytic approach. *Performance Evaluation*, 16(1–3):5–20, 1992.